

--2--

REMARKS

Claims 1-11 and 14-21 are pending¹. In view of the following Remarks, Applicants respectfully request reconsideration and timely withdrawal of the pending rejections for the reasons discussed below.

35 U.S.C. §103 Rejection

Claims 1-11 and 14-21 are rejected under 35 U.S.C. §103(a) as being unpatentable over Radev et al. ("Radev") (Generating Natural Language Summaries from Multiple On-line Summaries) in view of Cohen ("Cohen") (Fast Effective Rule Induction). This rejection is respectfully traversed.

In order to reject a claim under 35 U.S.C. §103(a), MPEP 2143, states, in part:

"To establish a *prima facie* case of obviousness,... there must be some suggestion or motivation, either in the references themselves or in knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine the reference teachings.... Finally, the prior art reference (or references when combined) must teach or suggest all of the claimed limitations."

Applicants submit that the references do not teach all the claimed limitations and that if the references were combined, the result would not provide the features of the invention.

The invention is directed to a method, apparatus and computer program product for providing summaries of documents belonging to a class of documents in a classified

¹ Claims 12 and 13 were not filed in the original application; hence, claims 12 and 13 do not exist.

document collection. In embodiments, a sample set of documents belonging to one or more classes is processed by a machine learning system in order to induce a set of rules associated with the sample set of documents. The words, phrases, and terms of an incoming document may be matched to the extracted set of rules or set of rules associated with the sample documents to provide a summary of each of the incoming documents.

Radev is directed to a system for filling predefined templates and then using a natural language generation techniques to provide a readable version of the formatted template, as verified by Figure 1 of Radev. As pointed out in the Background section of the invention (p. 3, lines 5-25), the technique of Radev is a common approach for automatically generating document summaries which relies on either natural language processing or quantitative analysis. Natural language processing is computationally intensive, while quantitative content analysis relies upon statistical properties of text to produce summaries. These types of approaches are resource inefficient as one of ordinary skill in the art would understand. The system of Radev is quite complicated and inherently dependent on very computationally intensive natural language analysis of each individual document for generating a summary.

Radev's reliance of templates can be seen in Figures 2 and 3. The invention, on the other hand, avoids reliance on templates and also avoids performing a resource intensive process on each individual document. Radev's approach is intrinsically resource intensive due to the natural language analysis and the templates involved for each individual document. However, Radev does not disclose inducing a set of rules from a sample set of documents and, contrary to the Examiner's assertion (i.e., section 1.1, pages 5-13), does not disclose comparing the extracted words, phrases and terms appearing in

the set of rules to an individual incoming document. The only mention of rules in Radev is on page 5, lines 8-16, where it is disclosed that new templates and information extraction rules for the templates would be required. Applicants note that the only reference to rules is in regards to the processing of templates. So, it appears that Radev contemplates rules for use of template creation/processing only and that the use of templates is still intrinsically maintained. This is contrary to the invention where the use of templates is not required or contemplated.

Cohen is directed to a rule pruning algorithm that performs efficiently on noisy sample datasets using reduced error pruning (REP) techniques to prune decision trees associated with noisy data. Cohen simply discloses creating and trimming rules. However, Cohen, similar to Radev, does not disclose any structure for, or steps of, comparing extracted words, phrases and terms appearing in the set of rules induced from sample documents to an individual input document.

Additionally, since Radev consistently discloses the requirement for using templates and filling those templates (e.g., Figures 1, 2 and 3 and page 5, lines 1-7) followed by processing the templates using natural language processing, then, if, *arguendo*, one were to combine the teachings of Radev and Cohen, the combination would not yield the features of the invention. At best, one would have a template-centric system for document summary production which may use tree-trimming algorithms to prune decision trees to generate rule sets. However, how the rule sets are used to generate summaries of documents is unknown. And, one of ordinary skill in the art would not know how to do this based on the Radev and Cohen disclosures. For example, the structure for, or steps of, comparing extracted words, phrases and terms appearing in the

set of rules induced from sample documents to an individual input document is never disclosed. This fictitious template-based system is not the same as the invention. The invention does not employ or contemplate the use of templates.

Since claims 1, 15 and 21 include one or more variations of these inventive aspects, e.g., a method or computer program component for inducing a set of rules from a sample set of documents and comparing extracted words, phrases and terms appearing in the set of rules induced from sample documents to an individual input document, or providing summaries based matches between extracted rules and an incoming document, these claims are also distinguishable.

As to claims 2-11, 13 and 14, Radev and Cohen do not teach or suggest, either singly or in combination, many of the features of the invention such as, for example, comparing the more than one individual incoming input document to the extracted words, phrases, terms appearing in the set of rules induced from the sample set of documents, comparing a same class of the at least two individual documents and the sample set of documents, or refining the incoming document including stemming, tokenizing, or morphological text processing.

Additionally, independent claim 16 is written in means-plus-function language such that, in order to reject such claim, the references must expressly or inherently perform a function identical to that of the means element, and the reference's structure for performing the function must be equivalent to that disclosed in the subject specification. *In re Donaldson Company, Inc.*, 16 F.3d 1189, 29 USPQ2d 1845 (Fed. Cir. 1994). MPEP §2182. However, the presently applied references do not show or suggest

an identical function, either expressly or inherently, and the structures of the applied references do not perform the equivalent functions of the claimed invention.

As to claim 16, Radev and Cohen do not show an identical function or equivalent structure as required by *In re Donaldson Company, Inc.* for (i) a “means for inducing a set of rules from a sample set of documents, the induced set of rules being characteristic of the sample documents,” (ii) a “ means for comparing extracted words, phrases and terms appearing in the set of rules induced from the sample set of documents to at least one individual incoming input document”, and (iii) a “means for providing a summary of each of the at least one individual incoming document based on matches between a vocabulary of the set of rules induced from the sample set of documents and the at least one individual incoming input document.”

As can be readily discerned, Cohen provides no structure at all since Cohen is directed to a mathematical algorithm for tree-pruning. In Radev, the structure provided is shown in Figure 1. Since Radev is directed to a natural language processing system using templates, there is neither equivalent structure nor identical function as required by *In re Donaldson*. The structure of Radev facilitates the templates and natural language processing. This is not the same as the present invention.

As to claims 17-20, Applicants submit that the identical function or equivalent structure are not disclosed or suggested by Radev or Cohen, either singly or in combination. For example, (i) a “means for extracting the sets of rules in order to provide a precise description of one or more classes associated with the sample documents”, (ii) “means for extracting a set of rules in order to provide a concise description of the one or more set of sample documents”, (iii) a summary of the

--7--

individual incoming document includes the word or term or phrase matches between the at least one individual incoming input document and the concise description of the extracted rules induced from the sample set of documents” (iv) a “ means for refining the individual incoming document... wherein the refining includes at least (i) stemming, (ii) tokenizing, or (iii) morphological text processing.

Applicants therefore submit that a *prima facie* case of obviousness has not been established and that the §103(a) rejections of claims 1-11 and 14-21 should now be withdrawn.

CONCLUSION

Applicants believe that a full and complete response has been made to the pending Office Action and respectfully submit that all of the stated grounds for rejection have been rendered moot. Accordingly, Applicants respectfully submit that all of the claims are allowable and that the application is in condition for allowance. Should the Examiner feel that there are any issues outstanding after consideration of this response, the Examiner is invited to contact the Applicants' undersigned representative.

Respectfully submitted,



Andrew M. Calderon
Reg. No. 38,093

McGuire Woods LLP
1750 Tysons Boulevard
Suite 1800
McLean, VA 22102-4215
Tel: 703-712-5426
Fax: 703-712-5285
\\COM378495.2